AWARD NUMBER: W81XWH-14-1-0345

TITLE: The Role of 5-Hydroxymethylcytosine in Gene Dysregulation of Epileptogenic Tubers in Tuberous Sclerosis Complex Patients

PRINCIPAL INVESTIGATOR: Dr. Alan Dombkowski

CONTRACTING ORGANIZATION: Wayne State University
Detroit, MI 48202

REPORT DATE: October 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2015	Annual	30 Sep 2014 - 29 Sep 2015

4. TITLE AND SUBTITLE

The Role of 5-Hydroxymethylcytosine in Gene Dysregulation of Epileptogenic Tubers in Tuberous Sclerosis Complex Patients

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0345

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Alan Dombkowski

E-Mail: domski@wayne.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Wayne State University
Carole Bach
5700 Cass Ave. Ste 4900
Detroit, MI 48202-3692

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Most tuberous sclerosis complex (TSC) patients suffer from epileptic seizures. For many patients anti-epileptic drugs do not control seizures. TSC seizures are associated with cortical malformations called tubers, but not all tubers cause seizures. It is unknown why some tubers cause seizures while others do not. We are investigating the hypothesis that 5-hydroxymethylcytosine (5hmC) in DNA contributes to dysregulation of epilepsy risk genes in epileptogenic TSC tubers. We are using semiconductor sequencing to characterize 5hmC in DNA of epileptogenic and non-epileptogenic brain tubers that were surgically resected from TSC patients to treat intractable epilepsy. During the reporting period we successfully implemented a method for base-resolution sequencing of 5hmC. We have also devised a bioinformatics strategy to enable comparison of genome-wide 5hmC profiles and identify regions of altered 5hmC abundance. Our preliminary results reveal localized genomic regions that have different 5hmC patterns in epileptogenic tubers compared to non-epileptogenic tubers and normal tissue. We are currently identifying such regions that are also proximal to known epilepsy risk genes. We will then measure gene expression levels for these genes. The outcomes of this research are expected to expand our understanding of the molecular contributors to epilepsy in TSC and identify novel therapeutic targets.

15. SUBJECT TERMS
Tuberous sclerosis complex, epilepsy, seizure, DNA methylation, hydroxymethylcytosine, 5hmC, next-generation sequencing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	14	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

Progress Report for Hypothesis Development Award #TS130067
The Role of 5-Hydroxymethylcytosine in Gene Dysregulation of Epileptogenic Tubers in Tuberous Sclerosis Complex Patients
October 29, 2015

1 INTRODUCTION:

Up to 90% of tuberous sclerosis complex (TSC) patients suffer from epileptic seizures, and many are refractory to pharmacotherapy. TSC seizures are usually associated with cortical malformations called tubers. However, not all tubers cause seizures. It is unknown why some tubers cause seizures (epileptogenic) while others do not (non-epileptogenic). We are investigating the novel hypothesis that 5-hydroxymethylcytosine (5hmC) in DNA contributes to dysregulation of epilepsy risk genes in epileptogenic TSC tubers. Discoveries in the past three years have revealed that 5hmC is very important in neural gene regulation. In this project we are using semiconductor sequencing (Ion Torrent) to characterize 5hmC in DNA of epileptogenic and non-epileptogenic brain tubers that were surgically resected from TSC patients to treat intractable epilepsy. We will then link the observed changes in 5hmC to dysregulated expression of genes known to confer risk of epilepsy. The outcomes of this research are expected to expand our understanding of the molecular contributors to epilepsy in TSC and identify novel therapeutic targets.

2 **KEYWORDS:** Tuberous sclerosis complex, epilepsy, seizure, DNA methylation, hydroxymethylcytosine, 5hmC, next-generation sequencing

3 ACCOMPLISHMENTS:

What were the major goals of the project?

Specific Aim 1: Identify 5hmC methylation patterns comparing epileptogenic and non-epileptogenic tubers. Months 1-12. 75% complete. The overall objective of this aim is to identify DNA 5hmC patterns that distinguish epileptogenic tubers from non-epileptogenic tubers and associate the 5hmC events with epilepsy risk genes. 5hmC profiling is accomplished using next-generation sequencing.

Specific Aim 2: Characterize differential expression of 5hmC-regulated epilepsy risk genes. Months 13-24. This effort will begin in Dec 2015. The overall objective of this aim is to characterize expression profiles for the epilepsy risk genes identified in Aim 1 as having altered 5hmC in epileptogenic tubers.

What was accomplished under these goals?

The first milestone for major task #1 is local IRB approval. We obtained the initial IRB approval in August 2014. A continuation of the IRB was approved in July 2015.

The second milestone for major task #1 is identification of epilepsy risk genes having differential 5hmC modifications. We have made excellent progress towards this milestone, and we are approximately 75% complete. We expect that we will complete this milestone by month 15 (Dec 2015). The original target

date for completion of this milestone was month 12. However, we made a number of improvements to our methods, implementing recently published techniques having greater specificity and coverage for 5hmC sequencing. This activity delayed the completion of the overall aim, but is providing excellent results. We anticipate that we will submit a first manuscript for publication detailing our sequencing and associated computational methods in month 16 (January 2016). A second manuscript is expected to follow in month 24, detailing 5hmC events associated with epilepsy. We believe that our approach to sequencing and analysis of 5hmC will be of considerable interest to investigators in a wide range of disciplines. Furthermore, we expect that the outcomes will have a significant impact on clinical application of 5hmC analysis, for example as biomarkers.

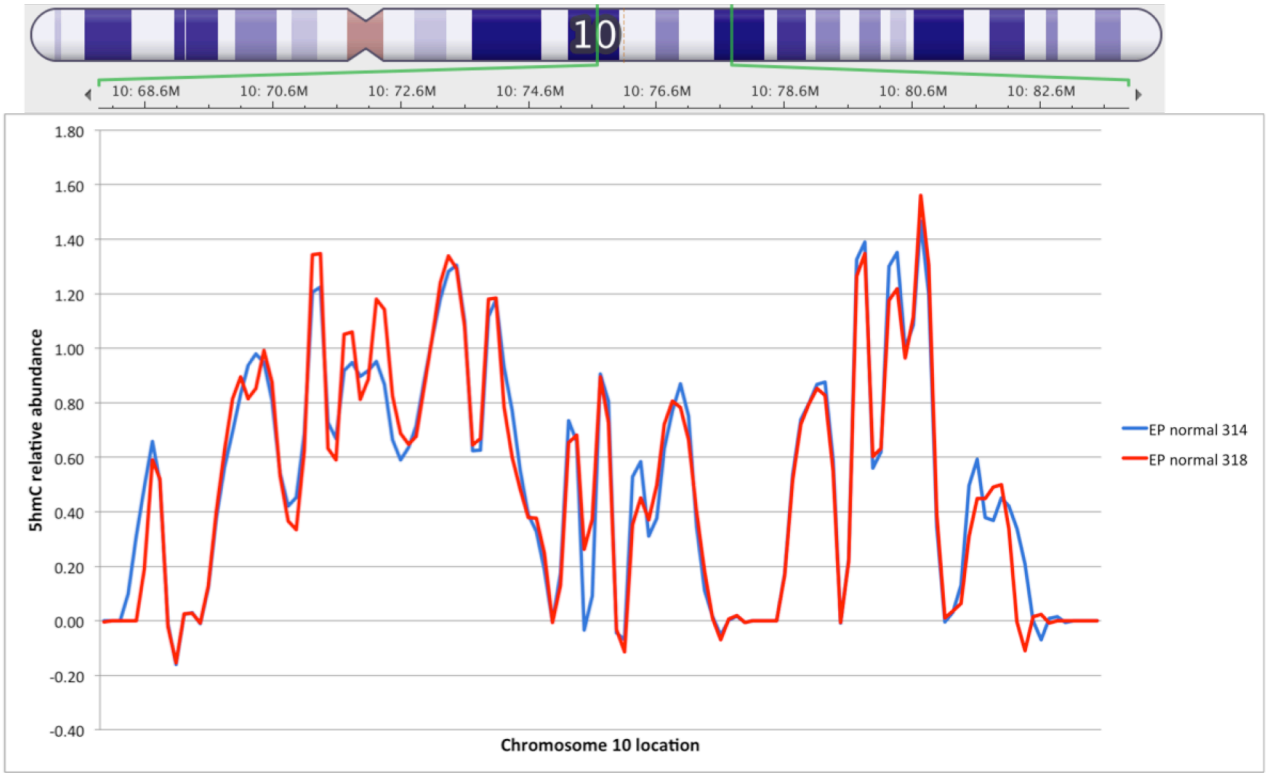

Figure 1. Comparison of 5hmC profiles from technical replicates demonstrates excellent concordance and robustness. Two independent sequencing runs were performed from a sample of normal human brain DNA. A representative region from chromosome 10 is shown. Importantly, the two sequencing runs were performed on chips with a 10X difference in sequencing capacity and output. Despite the 10X difference in reads, we are able to identify the same 5hmC profile.

In brief, sequencing of 5hmC is performed by first fragmenting DNA from a sample, then capturing fragments having 5hmC by using affinity-based methods, then constructing a library for sequencing by ligating technology-specific adapters onto the fragments. Efficient capture and enrichment of DNA fragments having a 5hmC modification is a critical step in 5hmC sequencing. Our initial results suggested that our original enrichment method introduced variability into the results. We have greatly improved our approach by incorporating improved techniques that were published after the submission of our proposal. Additional details are below in section 5 (Changes/Problems). Our initial sequencing during

months 3-6 (Dec 2014 – Mar 2015) indicated that considerable noise was inherent in the method, as highlighted by variation between technical replicates. Therefore, we implemented an approach published by another group in February 2015 (Sun, Dai et al. 2015). This method (Pvu-Seal-seq) utilizes two steps to isolate 5hmC fragments, first using a restriction enzyme (PvuRts1I) that specifically cuts 11 or 12 bases from a 5hmC site, followed by affinity-based enrichment. In addition to its excellent specificity, the restriction cut allows identification of the specific base modified with 5hmC and the strand. The authors of this study reported significant improvements over TAB-seq in specificity, sensitivity, genome-wide coverage, and cost-effectiveness. Importantly, the newer method can be used with low amounts of starting material, a constraint that limits bisulfite-based methods. This feature is of great benefit when studying archived samples where only a small amount of tissue may be available. The Sun study was accomplished using the Illumina sequencing platform, but after correspondence with the authors we were confident that we could successfully modify the protocol for our Ion Torrent platform. During the period of May-July of 2015 we performed the requisite modifications and testing of the modified Ion Torrent method. This work included the design of custom adapters specific for the Ion Torrent system, optimization of DNA fragmentation, and the development of positive control plasmid DNA. Our initial results with the modified approach were very promising, so we proceeded with preparation and sequencing of our target samples. To date, we have extracted DNA and RNA from 16 of the 20 target samples. In Aug-Oct 2015 (months 11-13) we prepared sequencing libraries for 13 samples using the new enrichment methods. We completed sequencing runs for 11 of these libraries. Preliminary analysis of the results is highly encouraging, and we expect that we will complete analysis and our associated milestone in month 15 (Dec. 2015).

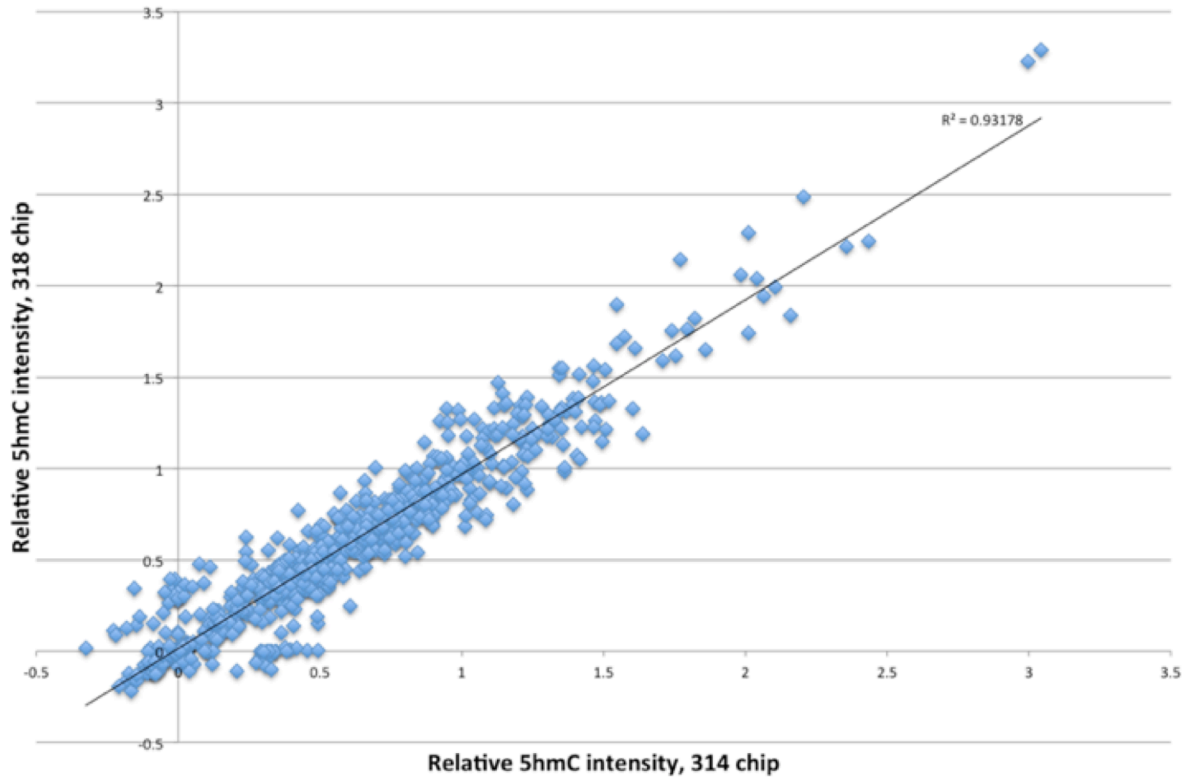

Figure 2. An excellent linear correlation is obtained for replicate sequencing runs. Each data point represents a region spanning ~ 130,000 bp in the region of chromosome 10 shown in Figure 1. The two axes represents relative 5hmC abundance measured for each region in two technical replicates with a 10X difference in sequencing capacity.

We consistently achieve ~one billion bases of high quality reads in each sequencing run using our modified approach (Pvu-Seal-seq) and the Ion Torrent 318 chip. We have devised a robust bioinformatics strategy to characterize 5hmC profiles in each sample and enable the identification of differentially populated 5hmC regions. This approach includes signal-processing techniques used in other disciplines. Our results demonstrate that our current sequencing and data analysis methods are reliable and reproducible. Comparison of data derived from technical replicates enables assessment of variation introduced by the assay. **Figure 1** shows the relative 5hmC density over a representative region of chromosome 10. Two curves are displayed, one for each of two sequencing runs derived from the same biological sample. In this case we used commercially available normal human adult brain tissue (Epigentek). We used this DNA extensively during our assay optimization to avoid depleting our patient samples. Importantly, the two replicate sequencing runs were performed on two different semiconductor chips having different capacities. Our Ion Torrent Personal Genome Machine is based on semiconductor sequencing and allows the use of several different chip capacities. Using the lowest capacity chip (314) we typically obtain 100 million bases of high quality reads. The highest density chip (318) provides 10 times the output, typically 1 billion bases. Despite a 10-fold difference in sequencing output of the two runs, we are able to measure highly concordant 5hmC patterns for the technical replicates. **Figure 2** demonstrates the strong linear correlation in measured 5hmC abundance when comparing the technical replicates. These results suggest that we may be able to use our analysis methods to integrate and compare 5hmC data that was derived from different sequencing platforms and having a range of output. This would be an additional outcome of our research with considerable impact. We expect to submit an initial manuscript detailing these methods in January 2016.

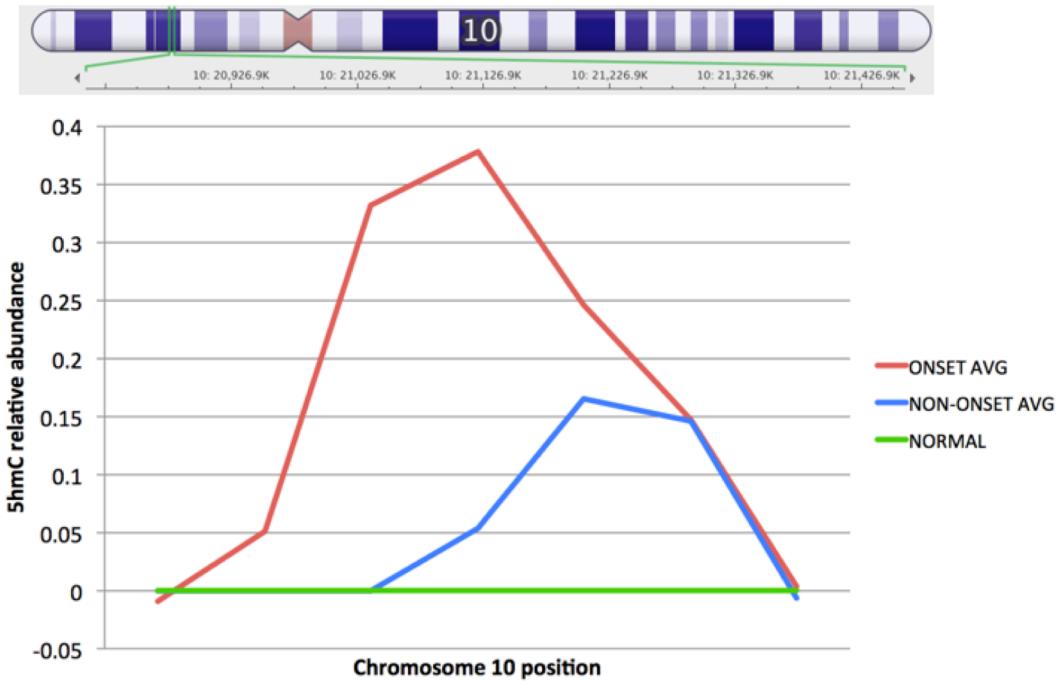

Figure 3. Example of a genomic region with altered 5hmC in seizure onset tubers. A 500k base region on chromosome 10 has elevated levels of 5hmC in seizure onset tubers (red), compared to non-onset tubers (blue) and normal control tissue (green). Each curve represents the average of 5hmC abundance in samples from multiple patients. This genomic region (10p12.32) has previously been implicated in epilepsy.

We are in the process of applying our computational strategy towards the identification of genomic regions with different 5hmC patterns when comparing epileptogenic (onset) tubers to non-onset tubers. Recent epilepsy surgeries in our institution have provided two valuable sample sets that include normal, non-tuber tissue in addition to onset and non-onset tissue for each patient. Occasionally during the course of surgery a small amount of non-tuber tissue is resected as part of a block resection. All three tissue types (normal, onset tuber, non-onset tuber) are rarely available for a single patient. These cases are very valuable because they allow for comparison across the three tissue types while controlling for numerous other factors such as medication, gender, age, etc. Our initial analysis of 5hmC profiles on these six samples, along with two other onset tubers from another patient reveal numerous regions with a difference in 5hmC between onset and non-onset tubers. An example is shown in **Figure 3**. This region spans approximately 500,000 bases on chromosome 10 and is located in a previously reported copy variant region associated with epilepsy (Striano, Coppola et al. 2012). Potential genes of interest in this region are PLXDC2, NEBL, and miR4675. The curves represent the average relative 5hmC density for four onset tubers (red), two non-onset tubers (blue), and two normal tissue samples (green). Our results indicate enrichment of 5hmC in the seizure onset tubers through this region. Our effort in the next two months will be to identify all regions of differential 5hmC that are proximal to genes that have previously been implicated in epilepsy. In the following eight months we will use expression profiling of these genes (Aim 2) to identify the associated epilepsy risk genes that are differentially expressed in epileptogenic tubers.

What opportunities for training and professional development has the project provided?

While this project is a hypothesis development grant and was not intended to provide training and development, we have been able to train one undergraduate student under this project, and we have recruited a second undergraduate student who will also serve as an intern on this project.

During the summer of 2015 Chae Kyung Jeon worked on this project as part of the Wayne State University Summer Undergraduate Research Program (SURE). Chae is a junior pre-med major at Grinnell College. During the SURE program Chae worked full time in our laboratory and learned all of the techniques required to extract DNA, isolate 5hmC fragments, create a sequencing library, and perform the sequencing run. Her efforts were instrumental in our successful implementation of the PvuRts enrichment method described above.

We have also recruited Ben Ciaglo, a senior Computer Science undergraduate at Wayne State University. Ben is highly interested in neurology and genomics, and he is preparing an application to the Wayne State Undergraduate Research Opportunity Program (UROP) to request a stipend for his effort on this project. We anticipate that Ben will start working part-time in our laboratory in Nov 2015 and will continue until August 2016. His efforts will be focused on developing computational pipelines to streamline our data analysis methods. He will be supervised by the PI (Dombkowski) and Co-I (Ghosh), and we anticipate that he will learn additional programming techniques, methods for clustering, and statistical analysis. We have already begun to teach Ben background information regarding tuberous sclerosis, epilepsy, and genomic technology.

How were the results disseminated to communities of interest?

We have not yet reported our findings; however, based on the encouraging results that we describe above, we expect to submit a manuscript for publication detailing our implementation of the PvuRts

method for the Ion Torrent system along with the associated computational techniques. Our goal is to submit this manuscript in January of 2016. A second manuscript detailing our complete findings associated with epilepsy risk genes is expected to be submitted near the end of the overall project period (Sept. 2016).

What do you plan to do during the next reporting period to accomplish the goals?

We are in the process of completing the remainder of effort in Aim 1, and we anticipate that the Aim 1 milestones will be met in December 2015. The remaining work entails completing the sequencing runs for nine samples, several of which already have libraries prepared. We anticipate all 5hmC sequencing will be completed by the end of Nov 2015. During this time our ongoing data analysis will continue as described above to identify epilepsy risk regions with altered 5hmC patterns in seizure onset tubers. The resulting list of genes will then be used in Aim 2 to determine associated gene expression profiles.

4 IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

Our initial results suggest that outcomes of this project will have a significant impact on tuberous sclerosis and epilepsy research, and contribute significantly towards methods for 5hmC analysis. We anticipate that two publications will result from this work, one detailing methods and one detailing the findings associated with tuberous sclerosis complex and epilepsy.

We are able to identify genomic regions with altered levels of 5hmC in epileptogenic tubers. When published, this will represent the first report of 5hmC in tuberous sclerosis and epilepsy. We will fully characterize such alterations across the genome and associate these events with changes in expression of known epilepsy risk genes. These results will enable us to formulate new hypotheses on epileptogenesis in tuberous sclerosis complex. Additionally, we are developing computational approaches for 5hmC analysis that we believe will enable integration of 5hmC results obtained from disparate sequencing platforms and capabilities. This would provide a considerable advance because it will facilitate the sharing of data and allow investigators to leverage results obtained in other laboratories.

What was the impact on other disciplines?

We believe that our bioinformatics approach will be very useful for 5hmC analysis in a wide range of disciplines.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

Nothing to report.

5 **CHANGES/PROBLEMS:**

Changes in approach and reasons for change

We did not have any changes to the scope or objectives; however, we did modify our experimental approach to incorporate recent advances in 5hmC profiling. We view these changes as very favorable to the overall quality of our data and the outcomes of this study. Furthermore, we expect that this effort will produce an additional manuscript detailing the methods, as we believe they will be of great use in a variety of biomedical studies. As reported above in the Accomplishments section, our initial sequencing results indicated that our original enrichment method might incur variability into the assay. Around that time a new method (Pvu-Seal-seq) for 5hmC enrichment and sequencing was published and offered significantly improved specificity, sensitivity, and genomic coverage (Sun, Dai et al. 2015). We allocated effort to adapt the Pvu-Seal method for our Ion Torrent sequencing platform. This effort included the development of adapters compatible with the Ion Torrent sequencing reaction, optimization of DNA fragmentation, and development of a positive control sample. The adapters are ligated onto the ends of each DNA fragment that contains 5hmC. One adapter must adhere to the "sticky end" created by the PvuRts1I restriction enzyme that cuts 11-12 bases away from 5hmC sites. We developed a positive control DNA that we use as a spike-in to assess 5hmC capture and enrichment. The control DNA is created by replicating DNA from a plasmid while incorporating 5hmC into the nucleotide mixture. These efforts proved successful, and our results indicate that the data are reproducible and robust (**Figures 1-2**). We have also expanded our list of epilepsy risk genes by including a recently published database (Ran, Li et al. 2015) and inclusion of published copy variants associated with epilepsy.

Actual or anticipated problems or delays and actions or plans to resolve them

As noted above, we required three months of effort to adapt and implement the Pvu-Seal-seq method for use with our Ion Torrent platform. This has delayed the completion of our 5hmC sequencing in Aim 1 and the start of Aim 2. However, we believe that the computational pipelines that we will develop with the assistance of our incoming Computer Science intern will expedite much of the analysis in Aim 2 and offset the delay.

Changes that had a significant impact on expenditures: None

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents: None

Significant changes in use or care of human subjects: None

Significant changes in use or care of vertebrate animals: None

Significant changes in use of biohazards and/or select agents: None

6 **PRODUCTS:** *List any products resulting from the project during the reporting period. If there is nothing to report under a particular item, state "Nothing to Report."*

Publications, conference papers, and presentations: Nothing to report.

Website(s) or other Internet site(s): Nothing to report.

Technologies or techniques:

We believe that our 5hmC sequencing and analysis approach will be of considerable interest to investigators in a range of biomedical disciplines. Therefore, we plan to publish details of the methods in January 2016.

Inventions, patent applications, and/or licenses: None

Other Products: None

7 **PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS**

What individuals have worked on the project?

Name:	Alan Dombkowski
Project Role:	Principal Investigator
Nearest person month worked:	3.5
Contribution to Project:	Dr. Dombkowski has overall responsibility for the project, including planning, experiment design, and supervision of assistants. Additionally, he is performing the initial bioinformatics processing and analysis of the 5hmC sequence data.
Funding Support:	WSU Department of Pediatrics. (Complete only if the funding support is provided from other than this award).
Name:	Dr. Harry Chugani
Project Role:	Co-Investigator
Nearest person month worked:	0.25
Contribution to Project:	Dr. Chugani coordinates the acquisition and classification of clinical specimens. He also provides expertise on interpreting results in the context of epilepsy and tuberous sclerosis complex.
Funding Support:	(Complete only if the funding support is provided from other than this award).

Name:	Samiran Ghosh
Project Role:	Co-Investigator
Nearest person month worked:	0.25
Contribution to Project:	Dr. Ghosh provides statistical expertise for experiment design and data analysis. His tasks include clustering of patient samples based on the 5hmC profiles.
Funding Support:	(Complete only if the funding support is provided from other than this award).
Name:	Daniela Cukovic
Project Role:	Research Assistant
Nearest person month worked:	4.5
Contribution to Project:	Daniela is responsible for molecular extractions, processing sample libraries, and performing the sequencing assays.
Funding Support:	WSU Department of Pediatrics (Complete only if the funding support is provided from other than this award).
Name:	Chae Kyung Jeon
Project Role:	Summer Intern
Nearest person month worked:	3
Contribution to Project:	Chae performed an internship under the Wayne State University Summer Undergraduate Research Program (SURE). She assisted with implementation of the Pvu-Seal-seq methods and development of the spike-in control.
Funding Support:	WSU Summer Undergraduate Research Program (Complete only if the funding support is provided from other than this award).

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Dr. Samiran Ghosh added the following support:
PCORI: IHS-1409-21410 (Patient-Centered Outcomes Research Institute*)
Title: "Developing Bayesian Methods for Non-inferiority Trial in Comparative Effectiveness Research."
Role: Principal-Investigator, Percent Effort: 37%, Period: 10/2015-9/2018

What other organizations were involved as partners?

Nothing to report.

8 SPECIAL REPORTING REQUIREMENTS

Nothing to report.

APPENDICES:

References Cited

Minetti and F. Zara (2012). "Clinical significance of rare copy number variations in epilepsy: a case-control survey using microarray-based comparative genomic hybridization." Arch Neurol **69**(3): 322-330.

Ran, X., J. Li, Q. Shao, H. Chen, Z. Lin, Z. S. Sun and J. Wu (2015). "EpilepsyGene: a genetic resource for genes and mutations related to epilepsy." Nucleic Acids Res **43**(Database issue): D893-899.

Striano, P., A. Coppola, R. Paravidino, M. Malacarne, S. Gimelli, A. Robbiano, M. Traverso, M. Pezzella, V. Belcastro, A. Bianchi, M. Elia, A. Falace, E. Gazzerro, E. Ferlazzo, E. Freri, R. Galasso, G. Gobbi, C. Molinatto, S. Cavani, O. Zuffardi, S. Striano, G. B. Ferrero, M. Silengo, M. L. Cavaliere, M. Benelli, A. Magi, M. Piccione, F. Dagna Bricarelli, D. A. Coviello, M. Fichera, C.

Sun, Z., N. Dai, J. G. Borgaro, A. Quimby, D. Sun, I. R. Correa, Jr., Y. Zheng, Z. Zhu and S. Guan (2015). "A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution." Mol Cell **57**(4): 750-761.